Int. J. Human–Computer Studies (1998) **48,** 521–545

Attention allocation within the abstraction hierarchy

MICHAEL E. JANZEN AND KIM J. VICENTE†

Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ont., Canada M5S 3G8. email: benfica@mie.utoronto.ca.

(Received 26 March 1997 and accepted in revised form 16 December 1997)

Previous research has shown that Rasmussen's abstraction hierarchy, which consists of both physical and functional system models, provides a useful basis for interface design for complex human–machine systems. However, very few studies have quantitatively analysed how people allocate their attention across levels of abstraction. This experiment investigated the relationship between attention allocation strategies and performance on a thermal-hydraulic process simulation. Subjects controlled the process during both normal and fault situations for about an hour per weekday for approximately one month. All subjects used a multi-level interface consisting of four separate windows, each representing a level of the abstraction hierarchy. Subjects who made more frequent use of functional levels of information exhibited more accurate system control under normal conditions, and more accurate diagnosis performance under fault trials. Moreover, subjects who made efficient use of functional information exhibited faster fault compensation times. In contrast, subjects who made infrequent or inefficient use of functional information exhibited poorer performance on both normal and fault trials. These results provide some initial, specific evidence of the advantages of an abstraction hierarchy interface over more traditional interfaces that emphasize physical rather than functional information.

1. Introduction

Rasmussen's (1985) abstraction hierarchy (AH) is a framework for representing the structure of complex human–machine systems. Each level of the hierarchy provides a different language for describing the system, with higher levels of abstraction containing functional information and lower levels containing physical information [see Bisantz & Vicente (1994) for a detailed example]. For process control systems, five levels of abstraction have been found to be of use:‡ (1) Goals—the goals for which the system was designed; (2) principles—the first principles (i.e. mass and energy conservation laws) describing system behaviour; (3) flows—the flow functions built into the system; (4) settings—the state (i.e. settings) of the equipment comprising the system and (5) form—the spatial location and visual appearance of the equipment.

There are two important reasons for representing human–machine systems in this way. First, because the AH describes the structure of the system to be controlled rather than the structure of human control tasks or actions, it provides operators with support

† Author to whom correspondence should be addressed.

‡ The labels used for the various levels of the AH were changed in this study in order to make it easier for subjects to understand the contents of each level. Rasmussen's (1985) original labels, from top to bottom, were functional purpose, abstract function, generalized function, physical function and physical form.

1071-5819/98/040521 + 25$25.00/0/hc970172

for events that are unfamiliar to then and that have not been anticipated by designers (Vicente & Rasmussen, 1992; Bisantz & Vicente, 1994). This is a crucial property since unanticipated events provide the greatest threat to the safety of complex systems (Vicente & Rasmussen, 1992). Second, because the AH is defined by means-ends relations between levels, it provides a psychologically relevant representation of a technical system supporting goal-directed problem solving (Vicente & Rasmussen, 1990, 1992). A number of verbal protocol studies in a variety of domains have shown that people naturally reason within an AH representation [see Vicente & Rasmussen (1992) for a review].

Given these important properties, it should not be surprising to find that several researchers have advocated that the AH be used as a basis for interface design (e.g. Goodstein, 1983; Vicente & Rasmussen, 1990, 1992). This approach involves conducting an AH analysis of a system to identify the information content and structure of the interface (e.g. Itoh, Sakuma & Monta, 1995; Dinadis & Vicente, 1996, in press). The resulting designs differ from traditional interfaces in that they provide both functional and physical information about the system, rather than primarily physical information [see Vicente, Christoffersen & Hunter (1996) for a detailed explanation].

There have been several attempts in industry, academia and government research laboratories to implement multilevel interfaces based on an AH representation [see Vicente (1992*a*, *b*) for a review]. More recently, controlled experiments have been conducted to evaluate the value of the AH for interface design in process control (e.g. Pawlak & Vicente, 1996; Christoffersen, Hunter & Vicente, 1996, 1997, in press), hypertext information search (Xu, 1996), and neonatal intensive-care units (Sharp, 1996). All of these studies have shown that an AH interface containing both physical and functional information can lead to better performance than a more traditional interface based on physical information alone.

Despite these demonstrated advantages, comparatively few quantitative data are available describing exactly how subjects allocate attention across the various levels of the AH. Vicente, Christoffersen and Pereklita (1995, Experiment 2) investigated this issue in a verbal protocol study and found positive statistically significant relationships between the use of functional information and fault diagnosis performance. However, their study was limited in a number of ways. First, the results were based on verbal protocol measures rather than more objective and comprehensive measures of attention allocation. Second, the simulation used in the study was not interactive, and as a result, subjects could only watch a real-time, dynamic trajectory of system behaviour rather than interactively control the system. Third, the subjects did not have any experience at controlling the system. Fourth and finally, the study focused primarily on fault situations rather than on normal situations. Consequently, there are comparatively few behavioural data on how experienced subjects allocate their attention across the various levels of the AH under normal and fault situations under real-time, interactive control conditions.

The study described in this article addresses this issue by investigating how subjects allocated their attention within the AH in a process control microworld. This, in turn, should allow us to determine which attention allocation strategies lead to enhanced performance. If the rationale behind the AH is correct (see Vicente & Rasmussen, 1992; Christoffersen *et al*., 1997), we should expect to find that subjects who demonstrate greater use of functional levels of information should exhibit better performance. An obvious way to test this hypothesis is to divide the content of an AH interface into multiple windows, each representing one level of the AH. By only allowing subjects to view one level at a time, we can then determine when and perhaps why subjects consult specific levels of abstraction.† Despite the apparent value of such a study, as far as we know, no experiment of this kind has ever been conducted.

2. System description

2.1. DURESS II

DURESS II is a real-time, interactive thermal-hydraulic process control microworld that has been designed to be *representative* (Brunswik, 1956) of complex work domains, thereby promoting generalizability of research results to operational settings (Vicente, 1991). The physical structure of DURESS II is illustrated in Figure 1(a). The system

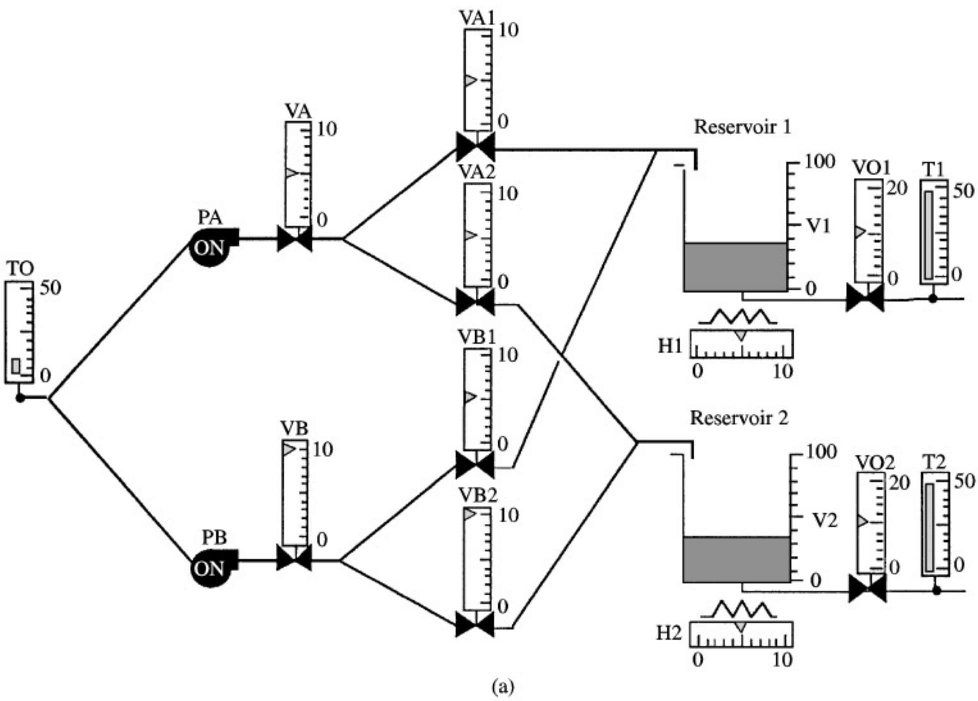
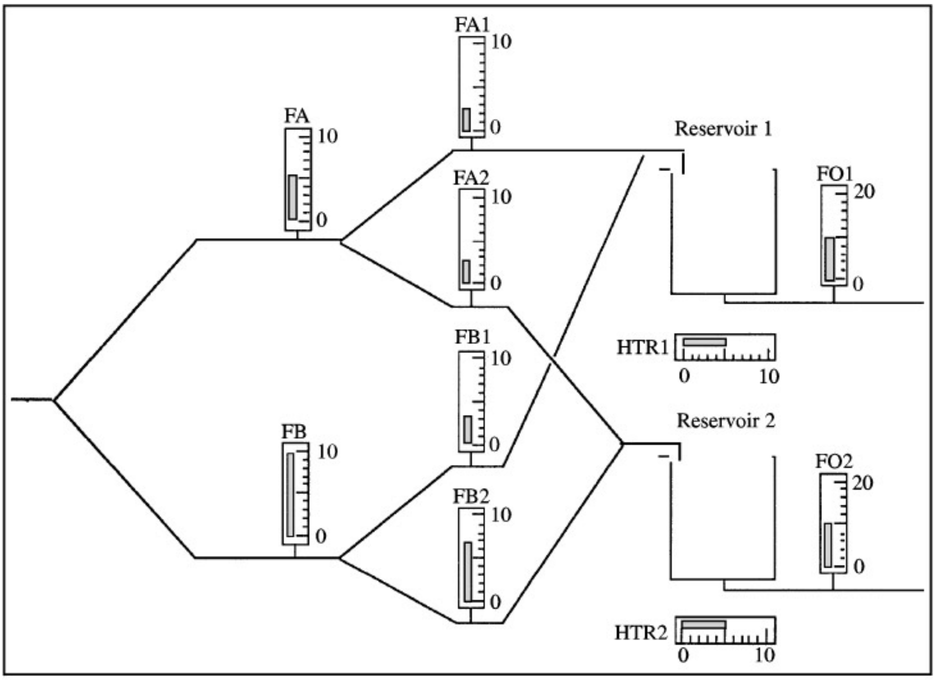
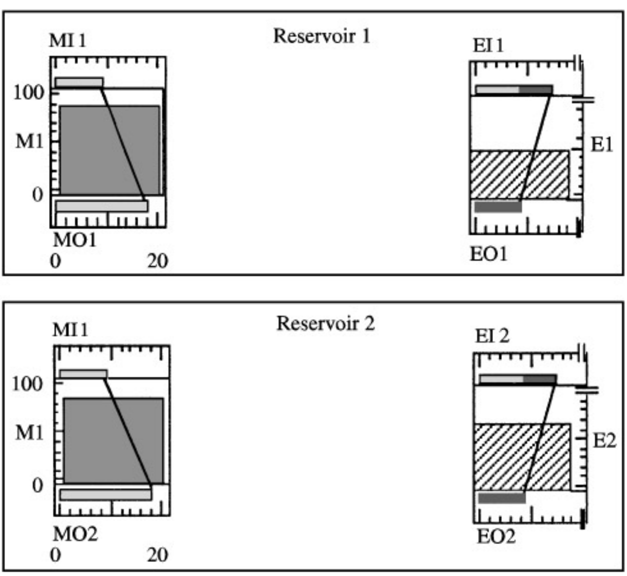


FIGURE 1. The Divided interface used in this study consisting of four independent windows, each representing one level of the abstraction hierarchy for DURESS II: (a) Settings; (b) Flows; (c) Principles; (d) Goals. See text for detailed description.

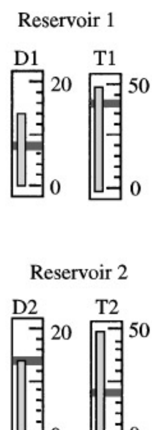
† This multi-window interface was designed to investigate empirically the relationship between attention allocation and performance. This interface is clearly not ideal, and is *not* intended to be a candidate design proposal to be generalized to industrial systems. For an example of how the ideas discussed here can be, and have been, implemented on a very large scale in industry, see Itoh *et al.* (1995).



(b)



(c)



(d)

FIGURE 1. (continued).

consists of two redundant feedwater streams (FWSs) that can be configured to supply water to either of two reservoirs. Each reservoir has associated with it an externally determined outflow 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 water output demands. 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 (H1 and H2). 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(a).

2.2. DURESS II ABSTRACTION HIERARCHY

An AH representation was created to describe DURESS II (Vicente & Rasmussen, 1990; Bisantz & Vicente, 1994).† As shown in Table 1, a part-whole dimension was used in addition to the means-ends dimension defined by the AH. Each shaded cell in the table represents a unique way of describing the DURESS II system (in much the same way that different maps give different information about the same geography). Only these cells contain meaningful representations of DURESS II; the other cells, which are not shaded, are not as useful.

2.2.1. Part-whole hierarchy

The part-whole hierarchy will be described first. Along this dimension, DURESS II can be viewed at three distinct levels of resolution: components, sub-systems and system. These three levels of the part-whole decomposition are column headings of the AH table (see Table 1). By moving from left to right, the system is decomposed into its smaller components. Conversely, by moving in the opposite direction, the components are aggregated to form the entire system. At the most detailed level, DURESS II is made up of different *components*: pumps, valves, reservoirs and heaters. *Sub-systems* are formed by aggregating these components into meaningful functional groups (e.g. the sub-system FWS A includes the components PA, VA, VA1 and VA2). This aggregation helps to reduce complexity by describing the system in less detail. Finally, if all sub-systems are combined, the entire *system* is viewed as a single entity.

2.2.2. Abstraction hierarchy

Whereas the part-whole hierarchy is concerned with describing DURESS II in terms of "chunks" of different sizes, the AH describes the system from different conceptual viewpoints. In terms of the map analogy presented earlier, each means-ends level provides a different type of map of the system. Each of the four levels will be briefly defined.

The first level, called *Goals*, describes DURESS II in terms of the goals for which the system was designed. The next level, *Principles*, describes the system in terms of the laws

† The bottom level of physical form can only be represented in a computer interface using a video link. Thus, it was not included in this study.

TABLE 1
Abstraction hierarchy for DURESS II

Means-end dimension	Part-whole dimension		
	System	Sub-system	Component
Goals			
Principles			
Flows			
Settings			

of conservation of mass and energy. This is done by describing mass and energy flows to and from each reservoir sub-system, as well as mass and energy inventories. In the level below Principles, *Flows*, DURESS II is described in terms of heat and water flows. This level describes the system in a more concrete fashion than the previous level. Furthermore, the heat and water flow level also differs from the previous level in that it is useful to describe the system at both the sub-system and component levels of the part-whole dimension (at this level, each FWS forms a sub-system, and each pump, heater or valve is a component). The fourth level of the AH, *Settings*, describes the settings or states of the individual components.

2.3. DIVIDED INTERFACE

An AH interface for DURESS II had already been developed and evaluated in previous research (Vicente & Rasmussen, 1990; Pawlak & Vicente, 1996; Christoffersen *et al*., 1996, 1997, in press). However, that interface presented multiple levels of the AH all on one integrated display, so it was difficult to measure comprehensively and objectively what information subjects were attending to at different times. Therefore, a new Divided interface was developed for this study by taking the information in the previous interface and dividing it into four windows, each corresponding to one level of the AH.† Subjects used a control panel with four buttons to select one of the four windows to be displayed at any time. These four windows are described next.

2.3.1. Settings

The Settings window in Figure 1(a) displays the state of all of the system components. 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. 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 grey (with black

† The allocation of system variables to each level of abstraction was also constrained so that each level would contain an independent set of variables. This made it easier for us to determine which information set subjects were consulting as they switched between windows.

lettering) if they are on. The next set of components are the primary valves (VA and VB), which have a continuous range from 0 to 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. The reservoirs have a maximum capacity of 100 units. Reservoir volumes (V1 and V2) are indicated by a scale on the side of each reservoir and by the shaded area depicting water in the reservoir. Below each reservoir is a heater (H1 and H2) that can be set to different heat transfer rates in a similar manner as valves are set. The only difference is that the heater setting indicators are red, whereas the valve setting indicators are yellow. To the right of each reservoir is an output valve (VO1 and VO2). The output valve scales have 2-unit increments, beginning at 0 and ending at 20. To the right of the output valves are thermometers (T1 and T2) which give the temperatures of the corresponding reservoirs. These thermometers are identical to the input thermometer.

2.3.2. Flows

The Flows window in Figure 1(b) displays, not the state of the components, but rather the state of the functions that those components are intended to satisfy. In the case of DURESS II, this corresponds to heat and water flows. Thus, each valve from the Settings window has been replaced by a flow meter in the Flows window. These meters show the flow rates of water through the respective valves (e.g. FA). Flows are shown in yellow and are measured using a scale from 0 to 10 units/s for FWS valves and from 0 to 20 units/s for the output valves (FO1 and FO2). Heat transfer rates (HTR1 and HTR2) from the heaters are indicated in a manner similar to the valve flow meters, the only difference being that the bars indicating heat flow are red.

2.3.3. Principles

The Principles window in Figure 1(c) displays additional higher-order functional information in the form of first principles (i.e. mass and energy conservation laws). The left-half of the graphic shows the mass balance. The meter on the top describes the flow of water into the reservoir (MI1 and MI2) and the one on the bottom indicates the rate of flow out (MO1 and MO2). The ends of these two flow rate bars are joined by a thin line. The slope of this line indicates the intended rate of change of volume. If the line has a positive slope (the input flow rate is greater than the output flow rate) then the mass should be increasing. On the other hand, if the slope is negative then there is a net outflow of mass, so the mass should be decreasing. When the line is vertical, then input equals output and the reservoir's mass should be constant. The varying slope of this line can be thought of as representing a funnel metaphor; when the funnel is wider at the top than the bottom (positive slope), the mass increases; when there is a negative slope, the funnel is inverted and the mass decreases. Mass inventory (M1 and M2) is measured using a scale on the side of the reservoir marked in 10-unit increments beginning at 0 and ending at 100.

The right side of the graphic in Figure 1(c) shows the energy balance for each reservoir, which functions in a similar manner. The energy input flow rate (EI1 and EI2) is shown at the top of the display. This value is broken down further into its two constituent parts: the light bar shows the rate at which energy is being contributed by the incoming water, while the dark part shows the rate at which energy is being added by the heater. The energy output flow rate (EO1 and EO2) is shown at the bottom of the display. As with the mass balance display, the line connecting the input and the output bars is based on a funnel metaphor and also represents the intended inventory gradient, this time of energy. The energy inventory (E1 and E2) itself is indicated by the scale on the side. For a description of how these displays were designed, see Vicente and Rasmussen (1990); for a detailed graphical description of these displays, see Pawlak and Vicente (1996).

2.3.4. Goals

The Goals window in Figure 1(d) displays the current state of the goal variables as well as their desired values, indicated by the shaded areas on the scales. Thus, the only information displayed is the temperature goals (T1 and T2) and the output demand goals (D1 and D2). There is a tolerance of $\pm 1.5^{\circ}$C around the temperature goals (40°C for reservoir 1 and 20°C for reservoir 2), and a tolerance of ± 1 unit/s on the outflow demand goals.

2.4. PURPOSE OF THIS STUDY

Three questions are addressed in this article: (1) How was attention allocated to the various levels of the AH? (2) How did attention allocation strategies differ between normal and fault trials? (3) Which attention allocation strategies were associated with superior performance?

3. Method

3.1. SUBJECTS

There were six subjects, primarily male engineering graduate students. Each subject was paid \$5 per session. An added bonus of \$2 per session was offered to subjects who completed the entire experiment. In an attempt to motivate subjects, a performance bonus was also offered. Neither the amount of this bonus nor the criterion upon which it would be based were specified. The value of the bonus amounted to approximately two additional dollars per session, per subject.

3.2. APPARATUS

Two Silicon Graphics IRIS Indigo R4400, Entry-Level Workstations were used to operate the DURESS II simulation. Verbal protocols were recorded using two Sony CCD-TR81 Hi-8 video cameras. All statistical data analyses were performed using SAS.

3.3. PROCEDURE

3.3.1. Introductory sessions

In the first session, subjects were presented with a description of the experiment, and a technical description of DURESS II, including a general overview describing the system structure and components. This description began with an introduction, followed by a statement of the objectives, a discussion of the material, self-assessment activities, a summary and a test to ensure that subjects had learned the requisite knowledge presented in the description.

In the second session, subjects were provided with an explanation of the interface, and a description of the types of trials that they would encounter throughout the experiment (see below). Any questions about the interface were answered by the experimenter, although questions about system functioning (i.e. the constraints governing the system) were not answered, as this was up to the subject to discover through experience with the system. Subjects were told that faults might occur and that, if they did, subjects should try to detect the fault, diagnose the root cause and compensate for it to achieve system objectives. Subjects did not know what types of faults could occur, when they would occur or how often they would occur. This was done to ensure that the fault management portion of the experiment was representative of actual process control systems.

In the third session, just before the first trial, subjects were provided with a description of the verbal protocol procedure that was to be used on the first trial and occasionally throughout the experiment. Subjects were required to give protocols on these initial normal trials so that they could gain experience speaking and controlling the system at the same time. Only protocols given during fault trials were actually recorded. A copy of all the forms and instructions that were used in this introductory session can be found in Howie, Janzen and Vicente (1996).

3.3.2. Task

Subjects were required to start up the system for each experimental trial. This required that the system be brought from a shut-down state, where all pumps, valves and heaters were off or closed and the reservoirs were empty, to a steady-state condition meeting the pre-defined temperature and outflow demand goals. A timer measuring the elapsed time in each trial was continually displayed in the upper left-hand corner of the screen. As a result, subjects could monitor the time they took to complete the various control tasks, if they so desired. For all trials, steady state was defined as having maintained the system goals (temperature and outflow demand) for both reservoirs for 5 consecutive minutes. If any one of the variables breached the goal tolerance, then the steady-state count was reset. All trials automatically terminated upon reaching steady state. Some trials required subjects to compensate for faults. If subjects had not achieved steady state within 30 min, the trial was terminated by the experimenter. If subjects performed actions which damaged system components (e.g. heating an empty reservoir, having a pump on when all downstream valves were closed) then the system "blew up" and the trial was automatically terminated. Subjects were required to control the system for about an hour per weekday for approximately one month. During this time, they performed a total of 67 trials. The trials and their order were identical for all subjects.

3.3.3. Trial types

There were three trial types: normal, routine fault and non-routine fault.

- *Normal trials.* During normal trials, there were no system anomalies. Output flow rate demand setpoints were varied between trials throughout the experiment, whereas the temperature goals remained the same from trial to trial (40°C for reservoir 1 and 20°C for reservoir 2).

- *Routine faults.* For routine fault trials, faults were chosen from a number of possible faults available in the simulation program, all of which are representative of faults that could occur in an actual process control plant. The routine faults were placed randomly throughout the experiment.

Routine faults were designed to be relatively simple in nature, and could be easily compensated for, once diagnosed. These were intended to be analogous to recurring failures in an industrial system, where some of the system components are inherently less reliable than others. Three classes of routine faults, occurring twice each, were used in this experiment.

- Valve blockages—one of the input valves became blocked, gradually eliminating flow.
- Heater faults—the output of one of the heaters was proportionately more (or less) than the setting.
- Reservoir leaks—one of the reservoirs developed a leak which drained the reservoir at a constant rate.

A total of seven routine fault trials were presented throughout the experiment. The seventh was a relatively simple fault where valve B2 became stuck at a setting of 10.

- *Non-routine faults.* Non-routine faults were intended to be analogous to rare, unanticipated occurrences within an industrial system. Two of the non-routine faults were placed towards the end of the experiment so that the effects of system experience could be evaluated, and the remaining one occurred towards the beginning of the experiment. The three non-routine faults given in the experiment were the following.
- A reservoir leak at 4 min into the trial, followed 2 min later by the addition of extra heat from an external heat source to the same reservoir.
- A simulated leak from reservoir 1 into reservoir 2 at 3 min into the trial.
- A heater fault where the heater's output became 150% of the setting at 3 min into the trial, followed by an increase of the incoming water temperature by 10°C 2 min later.

Note that all three scenarios included multiple symptoms which interacted (e.g. the heater fault and the temperature increase both affect the energy balance), making them more difficult to diagnose and compensate for.

3.1.4. Performance measures

There were two primary sources of data: simulation logs and verbal protocols.

- *Simulation logs.* The simulation logged all of the subjects' control actions and window transitions, as well as the state of all of the system variables. These raw data could be transformed in a number of ways to derive higher-order measures of performance.
- *Verbal protocols.* Verbal protocols were collected on video tape. Subjects were instructed to verbalize their thought processes, allowing the experimenters to hear what the subjects were saying while simultaneously observing the system state through the interface the subjects were using. Protocols were required on most fault days but also on some non-fault days so that the protocol was not a cue that a fault would occur. For

the same reason, one routine fault trial was not videotaped. Because verbal protocols were not collected for that trial, detection times and diagnosis scores are not available.

4. Results and discussion

The results are divided into three sections, each with its own discussion section. First, differences in attention allocation across levels of the AH and across individuals based on ANOVA are presented. Second, the correlation between attention allocation strategies and various measures of performance using regression analyses are presented. Third, individual differences between subjects are explored in a qualitative, integrative fashion in more detail. For additional analyses of the data from this study, see Howie *et al.* (1996).

4.1. EFFECTS ON ATTENTION ALLOCATION

Four measures of attention allocation were derived from the log files to determine how subjects' attention was distributed among the four windows: Frequency of Visits, Dwell Time, Relative Frequency of Visits and Relative Dwell Time. Data from trials that ended prematurely (i.e. "blowups") were excluded from the analyses. All *post hoc* comparisons were conducted with Student–Newman–Keuls (SNK) tests at $\alpha = 0.05$.

4.1.1. Frequency of visits

The Frequency of Visits was calculated as the average number of visits per min to each window for each trial for each subject. A four-way ANOVA was conducted with Window (Settings, Flows, Principles, Goals), Type (Normal, Routine Fault, Non-routine Fault) and Trial (number nested within Type) as within subjects-variables and Subject as a between-subjects variable. The results, shown in Table 2, reveal that all effects and interactions were statistically significant except for Type. Thus, Frequency of Visits differed across subjects, windows, trials and trial types. The Trial effects are confounded by learning and idiosyncratic differences in trial difficulty, and thus will not be analysed in more detail. The main effect and interactions involving Subjects are important because

TABLE 2
ANOVA results for frequency of visits

Source	df	F	p
Window	3, 15	19.98	0.0001
Subject	5, 291	22.18	0.0001
Window * Subject	15, 873	21.70	0.0001
Type	2, 10	1.56	0.2576
Window * Type	6, 30	3.29	0.0132
Subject * Type	10, 291	1.99	0.0343
Window * Subject * Type	30, 873	1.73	0.0091
Trial (Type)	63, 291	2.91	0.0001
Window * Trial (Type)	189, 873	2.42	0.0001
Subject * Trial (Type)	291, 873	3.31	0.0001

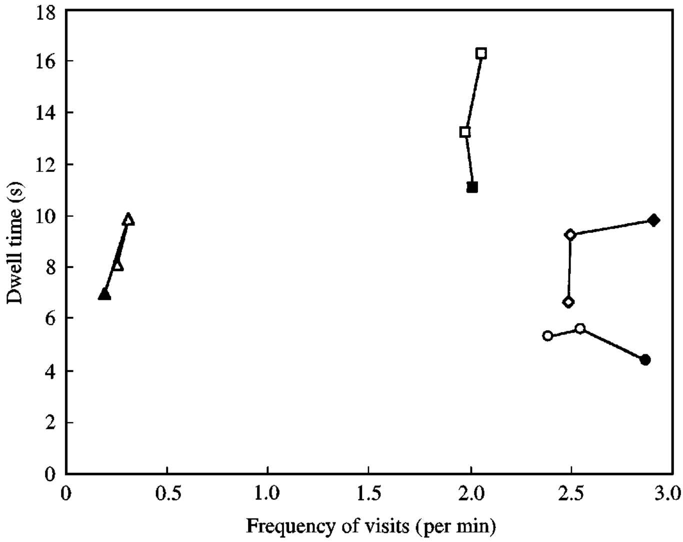


FIGURE 2. Mean attention allocation (Dwell Time and Frequency of Visits), as a function of Window (—□— Settings; —△— Flows; —○— Principles and —◇— Goals) and Trial Type (Normal, Routine Faults and Non-Routine Faults). For each window, the black marker represents the group means for all Normal trials; the markers connected to the black markers represent the group means for Routine Fault trials and the third markers represent the group means for Non-Routine Fault trials.

they reveal significant individual differences in attention allocation strategies. These will be discussed in later sections. The interaction between Window and Type is plotted in Figure 2 to illustrate how attention allocation to each window varied from normal trials to the two types of fault trials. For each window, the black marker represents the group means for all normal trials; the markers connected to the black markers represent the group means for routine fault trials and the third markers represent the group means for non-routine fault trials. Several patterns can be observed. First, the SNK analysis showed that Goals and Principles were visited significantly more often than Settings, which was visited significantly more often than Flows. Second, the impact of trial Type was strongest for Goals and Principles. During fault trials, these two levels were visited less often than during normal trials.

4.1.2. Dwell time

Dwell Time was calculated as the average time spent per visit to each window for each trial for each subject. A four-way ANOVA was conducted with Window (Settings, Flows, Principles, Goals), Type (Normal, Routine Fault, Non-routine Fault) and Trial (number nested within Type) as within subjects-variables and Subject as a between-subjects variable. The results, shown in Table 3, again reveal that all effects and interactions were significant except for Type. As with the previous analysis, the Trial effects will not be analysed in greater detail and the detailed analysis of individual differences will be deferred until later. Figure 2 shows the interaction between Window and trial Type on Dwell Time. For the Window effect, the SNK analysis showed that Settings was visited

TABLE 3
ANOVA results for dwell time

Source	df	F	p
Window	3, 15	13.66	0.0001
Subject	5, 291	13.10	0.0001
Window * Subject	15, 706	22.31	0.0001
Type	2, 10	1.85	0.2077
Window * Type	6, 28	4.81	0.0017
Subject * Type	10, 291	1.95	0.0383
Window * Subject * Type	28, 706	1.57	0.0322
Trial (Type)	63, 291	3.73	0.0001
Window * Trial (Type)	189, 706	1.54	0.0001
Subject * Trial (Type)	291, 706	1.70	0.0001

significantly longer than Goals, which was visited significantly longer than Flows, which was visited significantly longer than Principles. Figure 2 also shows that subjects dwelled longer for fault trials than for normal trials in all windows (especially Settings), except for Goals, in which subjects dwelled more briefly.

4.1.3. Relative frequency of visits

Because there were strong individual differences between subjects in terms of the total number of Frequency of Visits per trial, we normalized the Frequency of Visits for each window for each subject by dividing by the overall Frequency for each subject and each trial. This value was then divided by 4 so that the expected value of Relative Frequency of Visits for each window was 1, under the null hypothesis that attention would be uniformly distributed across windows. This normalization process filtered out individual differences in total frequency of visits and thereby provided an opportunity for gaining additional insight into each subject's attention allocation strategies. Because subjects' strategies changed over the course of the experiment, only data from the last 10 normal trials were used, to obtain a view of relatively stable, skilled strategies.

For each window, a one-way ANOVA was conducted on Relative Frequency of Visits, with Subject as the factor. For each window, there was a significant Subject effect [all $F(5, 54) \geq 26.31$, $p < 0.0001$], indicating that subjects differed in how often they visited each of the windows. The relationship between these individual differences in attention allocation strategies and performance will be addressed in detail later.

4.1.4. Relative dwell time

We also conducted analyses on normalized Dwell Time. For each window, Relative Dwell Time was calculated as the ratio of Dwell Time in that window to overall Dwell Time for the trial for each subject. The expected value of Relative Dwell Time in each window, assuming uniform attention allocation, was 1. Again, only the last 10 normal trials were analysed.

For each window, a one-way ANOVA was conducted with Subject as the factor. For Settings, Principles and Goals, there was a significant Subject effect [all

TABLE 4
Percentage of trials where flows was visited, by type and subject

Type	BL	CL	JC	RE	TN	YL	Average
Normal	100%	29%	93%	30%	36%	34%	54%
Routine fault	100%	40%	100%	57%	20%	60%	63%
Non-routine	100%	0%	100%	100%	0%	67%	61%
All trials	100%	29%	94%	35%	33%	38%	55%

$F(5, 54) \geq 16.73$, $p = 0.0001$]. The analysis for Flows is based on a smaller sample because not all subjects visited Flows on every trial, thereby creating missing data points. Despite the reduced power, the Subject effect was significant ($F(5, 31) = 9.90$, $p = 0.0001$), indicating that subjects differed in how long they visited each window. The relationship between these individual differences in attention allocation strategies and performance will be addressed in the following sections.

4.1.5. Visits to flows

Unlike the other windows, many subjects often never visited Flows during an entire trial. It is worthwhile examining these differences in more detail since they have implications for performance, as later analyses will show. Table 4 shows the percentage of all trials where Flows was consulted at least once, by subject and trial type. BL visited Flows on every trial, and JC almost always did. The other four subjects visited Flows for only 29–36% of normal trials, a marked difference. Of these four, RE and YL were more likely to visit Flows for fault trials. TN was less likely to visit Flows, the more complex the trial type was. CL was more likely to visit Flows for routine fault trials, but less likely to visit Flows for non-routine faults.

4.1.6. Discussion

Three important points emerge from these analyses. First, subjects allocated their attention in a non-uniform way across the various windows in the AH. In general, subjects looked most frequently at Principles and Goals. There seem to be several reasons for this. Monitoring Principles [see Figure 1(c)] is important because it allows subjects to quickly assess the stability of the system at a glance by looking at the sloped lines representing the expected mass and energy gradients. To the extent that these lines are close to vertical, the system should be stable. Monitoring Goals [see Figure 1(d)] is also important since it is the only way for subjects to monitor their progress with respect to system goals. This level thereby provides subjects with the information they need to determine that they have entered the goal region, thereby telling them that it is time to execute actions to stabilize the system in the goal area. Once they are in the goal area, this level also provides subjects with information to tell them that they are close to going outside the goal regions and thus that compensatory actions are in order. Finally, both Principles and Goals have the added advantage that they are relatively easy to perceive. Lower levels provide a more visually complex representation of the state of the system.†

Subjects also visited Settings [see Figure 1(a)] quite frequently, although less than Principles and Goals. This is most likely a result of the fact that this is the only level from which control actions can be implemented. Thus, any time subjects wanted to act on the system, they had to go to the Settings level. In addition, the Settings level also has an informational value, reminding subjects of the current settings of the components.

Interestingly, of all levels, Flows [see Figure 1(b)] was consulted least frequently. In fact, some subjects never consulted Flows at all during most of their trials (Table 4). The reason for this is that, under normal trials, the information provided in Flows is noticeably redundant with that in Settings. More specifically, if one knows the component settings (e.g. valve settings), it is a relatively simple matter to derive the state of the component functions (e.g. flow rates through the valves). This can be done with some relatively simple mental arithmetic, assuming that subjects know the correct relationships and that there are no faults in the system. Thus, the lack of emphasis on Flows during normal trials is understandable, given this redundancy.

The dwell times across levels of the AH were also non-uniform. Subjects dwelled the shortest in Principles, probably because of the emergent features in the mass and energy displays. The gradient slopes [see Figure 1(c)] make it possible for subjects to extract valuable goal-relevant information very quickly by using their powerful pattern recognition capabilities. Subjects dwelled the longest in Settings. This probably results from the fact that subjects usually made several consecutive control actions when starting up the system near the beginning of a trial. Thus, they would spend a fair amount of time, uninterrupted, in Settings. Subjects also dwelled fairly long in Goals because of the monitoring strategy they tended to adopt towards the end of the trial. Once subjects had stabilized the four goal variables in the desired region, they usually spent almost all of the remainder of the trial time in Goals where they just made sure that they stayed in the Goal region. In summary, these results indicate that there are good reasons for why subjects did not allocate their attention uniformly across levels of the AH.

A second finding to emerge from these analyses was that, in some cases, the subjects' attention allocation policies changed for faults. During these trials, subjects shifted more attention to lower levels of the AH. For example, subjects dwelled longer in Settings than for normal trials, in order to make more control actions that were required to compensate for faults. As well, some subjects were more inclined to visit Flows and to dwell longer there than they did for normal trials. This change in strategy is adaptive since the information in Flows is very useful for fault management. For example, a blocked valve fault (e.g. VA1) will cause the flow through that valve (e.g. FA1) to go to zero. On average, subjects visited Goals more briefly during fault trials than normal trials. The

† Visual complexity should not be equated with number of variables. This distinction can be observed by counting the number of variables presented in each window: Goals = 8 (including setpoints); Principles = 18 (including gradients); Flows = 10; Settings = 17. Although the Settings window looks much more complicated than the Principles window, it actually has a slightly fewer number of variables. Because the Principles level has emergent properties (i.e. conservation laws), an emergent features display (Bennett & Flach, 1992) could be constructed to integrate several variables into a single visual form that is easy to perceive. The Settings level of the AH does not have such emergent properties, making it difficult to design an emergent features display. As a result, variables have to be displayed in a separated fashion independently of each other.

aforementioned strategy of waiting in Goals at the end of normal trials after entering the goal region was no longer adopted by subjects. After a fault, subjects seemed to be more on their guard and wanted to monitor the system at lower (i.e. more detailed) levels to make sure that they had indeed compensated effectively for the fault and they could quickly detect any potential additional disturbances (which occurred in non-routine fault trials).

It is interesting to note that the attention allocation effects observed during a fault trial were similar to those which would be observed if the subjects had forgotten much of what they had learned and had had to start near the beginning of the experiment again. A detailed analysis of skill acquisition effects indicated that, with practice, subjects visited Principles and Goals more often and dwelled more briefly in Settings and Principles (Howie *et al*., 1996). Also, performance naturally improved. When a fault occurred, all of this was reversed. Subjects reverted to a more exploratory strategy. System behaviour was not as well understood, control was less accurate and efficient, and performance was poorer.

The third finding to emerge from these analyses of attention allocation strategies was that there were strong individual differences between subjects. The strongest differences was observed for the Flows window, which some subjects consulted on almost every trial and other subjects rarely consulted at all (Table 4). Are these effects idiosyncratic or do they have some relationship to performance?

4.2. ATTENTION ALLOCATION EFFECTS ON SKILL

This section describes the results of four multiple, linear regression analyses that sought to determine which attention allocation measures were good predictors of the subjects' control performance in normal or fault trials. One regression analysis was conducted for each of the following performance measures: trial completion on normal trials, fault detection time, fault diagnosis accuracy and fault compensation time. A total of 14 different attention allocation measures were used as predictors in each regression analysis. For each window, the following measures were included: Frequency of Visits, Relative Frequency of Visits, Dwell Time and Relative Dwell Time (because of missing data, Dwell Time and Relative Dwell Time for Flows were omitted). Only variables significant at the $p < 0.05$ level were entered into the regressions. All R^2 values refer to cumulative model R^2. We were interested in investigating whether these relationships existed rather than identifying quantitative models, so we have not reported the coefficients in the regression equations.

4.2.1. Normal trial completion time

Trial completion times represented the time that subjects took to bring the system from a shut-down state to a steady-state condition whereby the system goals were satisfied for five consecutive minutes. Trial completion times for blow-ups were counted as missing data. For time-outs, 30 min was recorded as a conservative estimate of trial completion time.

To study the relationship between attention allocation and asymptotic performance on normal trials, a stepwise linear regression was performed with data from the last 10 normal trials. Only one variable was entered into the model:

- The (negative) predictor entered was Relative Frequency of Visits to Principles [F(1,58) = 15.77, $R^2 = 0.21$, $p = 0.0002$].

The more often subjects visited Principles, the faster their performance on normal trials.

4.2.2. Fault detection time

To study the relationship between attention allocation and fault detection performance, a stepwise linear regression for fault detection time was performed for all fault trials. Detection Time was the time from the onset of a fault until subjects verbally reported detecting a fault (Pawlak & Vicente, 1996). For trials in which subjects did not detect the fault, Detection Time was counted as missing data. The predictor variable data (i.e. the attention allocation measures) were based only on the corresponding fault trials. No variables were entered into the model, indicating an absence of a significant linear relationship between attention allocation strategies and fault detection time.

4.2.3. Fault diagnosis accuracy

To study the relationship between attention allocation and fault diagnosis performance, a stepwise linear regression for fault diagnosis accuracy was performed for all fault trials. Diagnosis accuracy was extracted from the verbal protocols based on a method used by Pawlak and Vicente (1996) and Christoffersen *et al.* (1997). Recall that subjects were instructed to diagnose the root cause of each fault. Accordingly, each verbalized attempt at diagnosis was assigned a numerical value which measured the extent to which subjects were able to reach this ideal. There were four different levels of response, scored as 0, 1, 2 or 3. These levels are the following.

- 0—The subject says nothing relevant to the fault or says nothing at all.
- 1—The subject states that the system is not behaving as expected and describes, at a general level, the symptoms of the fault (e.g. "the level of reservoir 1 is dropping").
- 2—The subject describes fault symptoms at a more functional level, but fails to localize the fault (e.g. "I'm losing flow somewhere").
- 3—The subject localizes the root cause of the fault (e.g. "heater 1 is not functioning properly").

Each successive level is closer to the ideal of an accurate root cause diagnosis. Scores were assigned independently by the two experimenters, and the few disagreements that occurred were resolved by discussion. The highest level of diagnosis made for a particular fault was recorded as the Diagnosis Score for that trial. For each routine fault, there was one diagnosis score. For each non-routine fault, there were two: one for the first fault and one for the second fault. Thus, there were a total of 12 simple faults for which diagnosis scores were recorded.

In order to create a model to predict diagnosis score, a stepwise linear regression was performed on all faults, using attention allocation measures from the corresponding fault trials. Because the diagnosis score may not be on an interval scale, R^2 values are not meaningful and thus are not reported. Nevertheless, F tests can still be meaningfully conducted to determine if there is a significant relationship between any of the predictor variables and the ordinal dependent variable.

Two variables were entered into the model.

- The first (positive) predictor entered was Relative Frequency of Visits to Flows [$F(1, 53) = 26.90$, $p = 0.0001$].
- The second (positive) predictor entered was Frequency of Visits to Principles [$F(2, 52) = 4.47$, $p = 0.04$].

The more often subjects visited Flows and Principles, the more accurate their diagnosis scores.

4.2.4. Fault compensation times

To study the relationship between attention allocation and fault compensation performance, a stepwise linear regression for trial completion time was performed for all fault trials. Trial was added as a predictor to see if the variance in times was due to idiosyncratic trial effects rather than systematic strategy differences. Type was also added (as a dummy variable) to distinguish non-routine from routine fault trials. The following three variables were entered into the model.

- The first (positive) predictor entered was Dwell Time in Principles [$F(1, 45) = 20.00$, $R^2 = 0.31$, $p = 0.0001$].
- The second (positive) predictor entered was fault Type [$F(2, 44) = 6.75$, $R^2 = 0.40$, $p = 0.013$].
- The third (negative) predictor entered was Trial [$F(3, 43) = 8.03$, $R^2 = 0.49$, $p = 0.007$].

Thus, dwelling briefly in Principles was related to shorter completion times. Also, non-routine faults took longer to complete, and trials occurring later in the experiment took less time.

4.2.5. Discussion

Despite the large number of potential predictors, a relatively simple pattern of results emerged from these regression analyses, emphasizing the significant role of functional levels of the AH. For normal trial completion times, Relative Frequency of Visits to Principles was the only significant predictor, accounting for 21% of the variance. The more often subjects visited Principles, the faster their performance on normal trials. This is probably because the vertical lines showing the rate of change of mass and energy provided a useful cue for stabilizing mass and energy, which in turn stabilized temperature. The sooner subjects effectively stabilized the system, the quicker they would be able to complete the trial.

For fault diagnosis accuracy, there were two significant predictors: Relative Frequency of Visits to Flows and Frequency of Visits to Principles. Subjects who visited Flows and Principles more frequently exhibited more accurate diagnoses. Thus, these two levels of the AH, which contained functional information, helped subjects to trace through the multiple levels of abstraction until the source of the fault was identified. The nature of this relationship can be illustrated through a few examples. For instance, a heater failure can be best diagnosed by comparing the heater setting presented in the Settings window with the heat transfer rate presented in the Flows window. The relationship that usually links these two variables is violated in the case of a heater failure, so the functional information in the Flows window is essential for diagnosis. As discussed earlier, a similar argument can be made for the case of a blocked valve failure. These examples indicate that functional information is important for fault diagnosis performance (cf. Christoffersen *et al*., 1997).

For fault compensation times, three predictors were significant: Dwell Time in Principles, fault Type, and Trial number, accounting for 31, 9 and 9% of the variance, respectively. The fault Type effect indicates that non-routine faults took longer to complete because they are more difficult than normal trails. The Trial effect indicates that trials occurring later in the experiment took less time to complete due to learning. The third, and most significant, effect indicates that subjects who dwelled less in Principles tended to compensate for faults more quickly, or conversely, that subjects who dwelled longer in Principles tended to take longer to compensate for faults. Recall that Frequency of Visits to Principles was a significant positive predictor of diagnosis accuracy. This suggests that there may be a trade-off to presenting functional information: subjects who visited Principles frequently and exhibited good diagnosis scores may have had to dwell longer in Principles in order to thoroughly assess the event, thereby exhibiting longer fault completion times. If this trade-off exists, then accurate diagnosis would have to come at the cost of long compensation times. A *post hoc* correlation analysis indicates that these two performance variables are not significantly related [$r(4) = 0.18$, n.s.; see Figure 3(a)]. Thus, there is no evidence of a trade-off to presenting functional information. Instead, the positive correlation between Dwell Time in Principles and fault trial completion time seems to be due to the graphic emergent features of the Principles display [Figure 1(c)]. Subjects who were able to make efficient use of the graphics dwelled more briefly in Principles, and thus, completed the fault trials more quickly than other subjects.

In summary, the regression analysis results indicate that various measures of attention allocation to functional levels of the AH (Flows or Principles) were the best predictors of performance. In each case, thorough or more efficient use of functional information was associated with better performance. In contrast, there was no evidence indicating a significant relationship, positive or negative, between the use of physical information and any of the performance measures collected on normal or fault trials.

4.3. INDIVIDUAL DIFFERENCES

An exploratory, qualitative analysis was conducted to provide additional insight into the individual differences identified earlier. The attention allocation variables that were significant predictors of performance (i.e. Relative Frequency of Visits to Principles, Relative Frequency of Visits to Flows, Frequency of Visits to Principles and Dwell Time in Principles) were used as frames of reference for examining how the subjects' strategies differed. Similarly, the four performance measures discussed in the previous section (i.e. trial completion on normal trials, fault detection time, fault diagnosis accuracy and fault compensation time) were used as frames of reference for examining how the subjects' performance differed. We then looked for patterns that could further explain the findings presented in previous sections.

This informal analysis revealed that subjects could be meaningfully grouped in a productive way using the following two performance measures.

- Normal trial completion time (for the last 10 normal trials).
- Diagnosis score (for all faults).

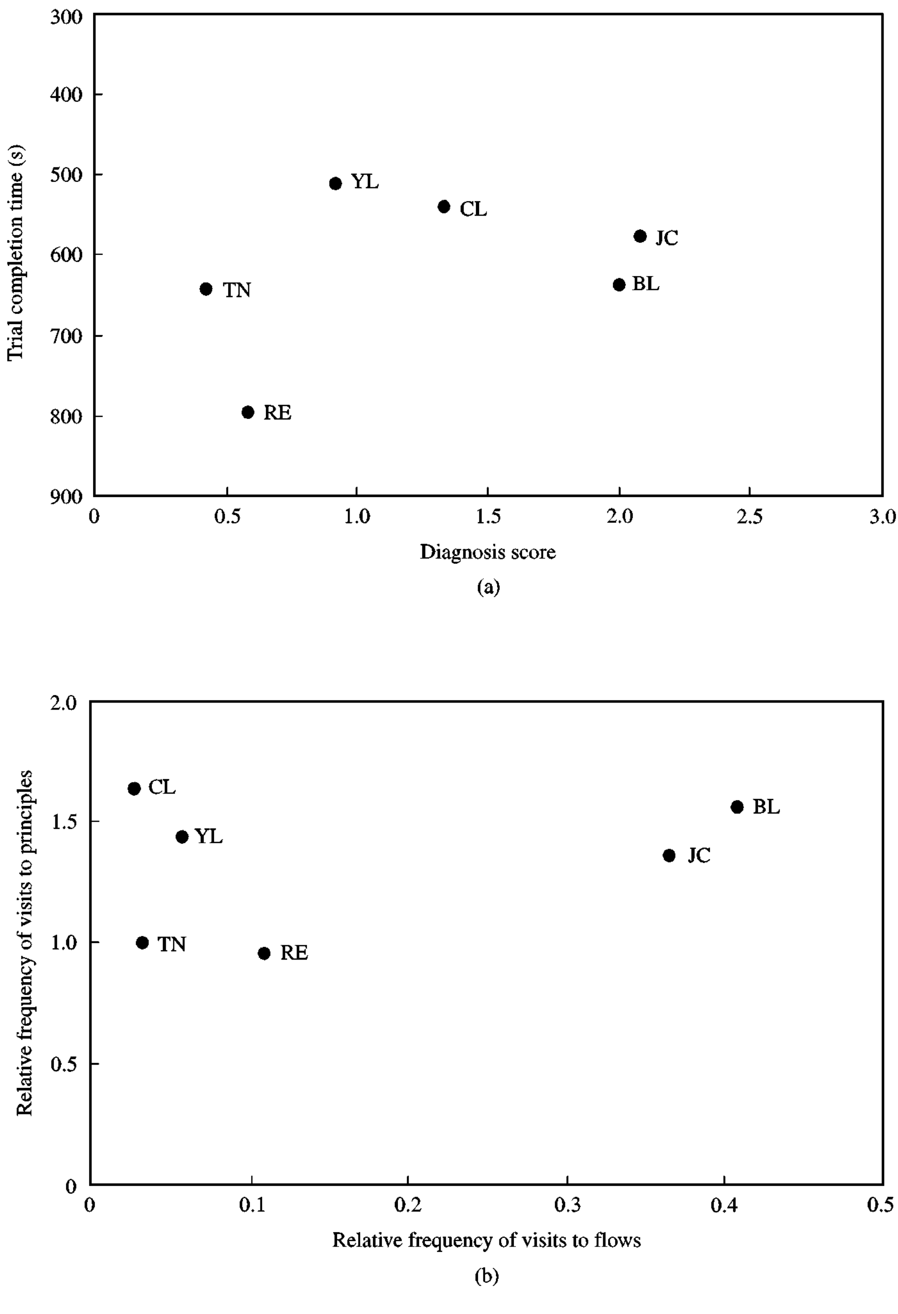


FIGURE 3. Individual differences between subjects in: (a) performance (normal trial completion time and fault diagnosis score) and (b) attention allocation strategies (relative frequency of visits to principles and relative frequency of visits to flows). Note that trial completion time on the y-axis is in reverse order, so that superior performance is towards the top of the graph.

Figure 3(a) shows each subject's mean performance on these measures. Note that Trial Completion Time on the *y*-axis is in reverse order, so that superior performance is towards the top of the graph. The two attention allocation measures that best predicted these two performance measures were the following.

- Relative frequency of visits to principles (for the last 10 normal trials).
- Relative frequency of visits to flows (for all fault trials)

Figure 3(b) shows the subjects' mean attention allocation strategies on these two measures.

By comparing Figures 3(a) and (b), it is apparent that the subjects' relative positions on the graphs are roughly similar, and can be categorized into the following three groups.

- Fast subjects (those with the fastest completion times): CL, YL.
- Insightful subjects (those with the highest diagnosis scores): BL, JC.
- Poor subjects (those with the slowest times and the lowest scores): RE, TN.

In the remainder of this subsection, these groups will be compared and contrasted. For simplicity and conciseness, passing reference will also be made to analyses not presented in this article to further explain the differences between groups [see Howie *et al.* (1996) for details].

4.3.1. Fast subjects

YL and CL had the fastest completion times at the end of the experiment. In addition to completing the startup task quickly, they also shared a number of other common features which further distinguished them from the other four subjects. Perhaps the strongest difference was that the Fast subjects visited Principles frequently but dwelled there briefly each time. The salient perceptual features offered by the mass and energy balance graphics [see Figure 1(c)] seemed to allow the Fast subjects to quickly judge whether mass and energy were rising or falling, and whether any adjustment was necessary to satisfy the Goals. Frequently consulting this information probably helped these subjects get the system in the goal region quickly, stabilize it there effectively and ensure that it stayed there for the required 5 min.

However, the Fast subjects rarely consulted the Flows window. As mentioned earlier, this was probably due to the fact that this information could be derived from Settings when the system was operating normally. The Fast subjects also dwelled long in the Goals window, which provides comparatively little information about the state of the system except whether the goals are achieved or not. This suggests that the Fast subjects were more concerned with meeting the goals than developing a comprehensive understanding of system state and functioning. Dwelling long in Goals allowed them to monitor their progress without distractions. Their priority appeared to be to achieve steady state, and any visits to windows below Goals were merely to support that objective.

The hypothesis that the Fast subjects emphasized speed to the detriment of system understanding is reinforced by the fault diagnosis results. During these trials, the Fast subjects made relatively few level 2 and level 3 diagnoses (compared to the Insightful subjects). On the few occasions when level 2 diagnoses were made, the Fast subjects were rarely consulting Principles (and never Flows). These findings suggests that the Fast subjects did not make full use of the functional information in order to make accurate diagnoses.

4.3.2. *Insightful subjects*

BL and JC were the best subjects at diagnosis. Interestingly, they began the experiment with the fastest average compensation times, faster than the Fast subjects in fact. The Insightful subjects continued to achieve very consistent times throughout the experiment. However, by the end of the study, they were somewhat slower than the Fast subjects.

The two Insightful subjects exhibited attention allocation strategies that distinguished them from the other four subjects. Perhaps the most unique feature was that they exhibited the most evenly balanced attention allocation strategies. Unlike other subjects who tended to ignore one or more windows, the Insightful subjects made use of all four windows. Figure 3(b) shows that they looked at Principles about as often as the Fast subjects, but that they looked at Flows much more often than any of the other subjects. This likely made them more familiar with the functional relationships governing the system, a hypothesis that is supported by their superior diagnosis scores. The Insightful subjects also spent more time than the Fast subjects in Settings and Principles, two windows that provide most of the information available from the interface. Collectively, this more even distribution of attention suggests that the Insightful subjects were interested, not just in finishing the task quickly, but also in monitoring the functional information to acquire a sound understanding of the system.

The advantages of the Insightful subjects' attention allocation strategy were revealed during fault trials. They made the most level 2 diagnoses of all subjects, almost always while looking at Flows or Principles. They also had the most level 3 diagnoses, which was partly a result of having the most level 2 diagnoses, since identifying the functional problem made it more likely to take the next step to infer the root cause. Since the Insightful subjects were in the habit of visiting Flows regularly during normal trials, they were also more likely to observe the faulty behaviour manifested in the functional information. Furthermore, they were more likely to correctly interpret that information, making it easier for them to trace the root cause and make a correct diagnosis.

4.3.3. *Poor subjects*

As shown in Figure 3(a), TN and RE were clearly the least proficient of all subjects, exhibiting the slowest normal trial completion times and the lowest diagnosis scores. It is important to understand how their attention allocation strategies differed from those of other subjects since this may help explain their poor performance with the very same interface.

The most distinctive feature shared by the Poor subjects was that they visited the functional windows (Principles and Flows) relatively less frequently than other subjects [see Figure 3(b)]. Poor subjects devoted most of their attention to Settings and Goals, effectively reducing their interface to be little more than a traditional interface containing only physical information. This offered a very restricted view of the system compared to that obtained by the Insightful subjects who divided their attention more uniformly across the four windows. This seems to be why Poor subjects performed poorly, much like the subjects in previous experiments who operated DURESS II using an interface with physical but no functional information (Pawlak & Vicente, 1996; Christoffersen *et al*., 1996, 1997, in press).

Interestingly, within this background of similarities, RE and TN differ from each other in the same qualitative way that the Insightful group differed from the Fast group, albeit on a smaller scale. Like the Fast subjects, TN rarely visited Flows (Table 4), and ironically, he was less likely to visit Flows the more difficult the trial type was. He visited Principles only briefly, which suggests that he placed a greater priority on speed than understanding, much like the Fast subjects. However, TN would rarely consult Principles to determine which direction the system was heading, and whether adjustments were necessary, so his strategy was less effective.

Like the Insightful subjects, RE visited Flows reasonably regularly, and dwelled long in Principles, showing an interest in functional information. It appeared that he placed a greater priority than TN on understanding the system. RE appropriately adjusted his attention to Flows for fault trials (Table 4). However, he did not visit Flows or Principles sufficiently often [Figure 3(b)] to achieve a reasonable level of performance [Figure 3(a)].

TN failed to make any level 2 or level 3 diagnoses, indicating that his understanding of the system was quite limited. RE made some level 2 diagnoses but only in Flows and Principles, indicating that he was making more effective use of the functional information than TN. Nevertheless, RE made very few level 2 diagnoses, and no level 3 diagnoses, indicating that he had acquired a limited understanding of the system compared to the Insightful subjects.

In summary, RE can be contrasted with TN in the same manner as the Insightful subjects can be contrasted with the Fast subjects. RE visited Flows more often and dwelled in Principles longer than TN. RE made level 2 diagnoses in the Flows and Principles windows, while TN did not. RE was better at diagnosis, while TN was faster. RE appeared more interested in understanding, while TN appeared more interested in speed. All of these patterns parallel the differences between the first two subgroups, although in a less pronounced manner.

5. Conclusions

The findings of this study show the value of using the AH as a basis for interface design for complex human–machine systems. To merely complete the task in this experiment, the only necessary windows are Settings and Goals. We know this because these two windows contain the state variables for the system. Furthermore, in the past, subjects completed the same task using an interface that had only these two levels of abstraction (e.g. Pawlak & Vicente, 1996; Christoffersen *et al*., 1996). Subjects who merely completed the task allocated most of their attention to one of these windows, and most of their remaining attention to the other. These were the least proficient performers. In order to not only complete the task, but to complete it quickly, it was necessary to be able to control the system accurately. This required frequent visits to Principles, which allowed reliable stabilization of mass and energy, which in turn stabilized temperature.

Subjects exhibiting this strategy were the fastest of all. To accurately diagnose faults, it was also necessary to be able to identify broken constraints. This was best accomplished by consulting the Flows levels defined by the AH. Subjects who adopted this strategy exhibited the best performance on fault trials.

This study also shows that interface design cannot guarantee effective performance. Not only is it necessary that goal-relevant information be available, but operators must be aware of the value of the information. Although all of our subjects had the very same information available, they used it quite differently, and with varying degrees of success. Without a suitable operator and/or proper training, a well-designed interface may be ineffective.

5.1. LIMITATIONS AND FUTURE RESEARCH

Despite these contributions, there are several limitations associated with this research, each of which leads to a future research topic. First, despite our attempts to achieve representativeness, DURESS II is much simpler than a real process control plant, and so attention allocation is not nearly as demanding as it is for industrial plants. The AH has been used to construct a multilevel interface for a full-scope nuclear power plant simulator (Itoh *et al.*, 1995) so we know that these design ideas can be applied to industry-scale systems (see also Dinadis & Vicente, 1996, in press). However, as far as we know, there are no data on how professional operators navigate through an AH-based interface to control a large-scale plant. It is important to conduct such a study to asses the generalizability of our findings. Second, the method by which attention allocation was measured in this study is clearly coarse in nature. We know what levels of the AH subjects were viewing, but not which particular variables. It would be useful to collect eye movement data to identify exactly what information subjects are looking at, at any one point in time. Although much more time consuming, this may allow us to understand the relationship between the current system state and the operator's cognitive processes on a dynamic basis (e.g. Moray & Rotenberg, 1989), rather than on an aggregate basis as in this study. Adopting a more sophisticated methodology would allow us to conduct a more refined investigation of how attention is allocated within the AH.

This research was sponsored by a contract from the Japan Atomic Energy Research Institute (Dr Fumiya Tanabe, Contract Monitor), as well as research and equipment grants from the Natural Sciences and Engineering Research Council of Canada. We would like to thank Dr Tanabe, Dianne Howie, Darryl Minard, Ian Spence and Thomas Smahel for their contributions. Also, thanks to Gunilla Sundström and three reviewers for their constructive comments.

References

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). *The Representative Design of Psychological Experiments*, 2nd edn. Berkeley, CA: University of California.

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.

CHRISTOFFERSEN, K., HUNTER, C. N. & VICENTE, K. J. (in press). A longitudinal study of the impact of ecological interface design on deep knowledge. *International Journal of Human–Computer Studies*.

DINADIS, N. & VICENTE, K. J. (1996). Ecological interface design for a power plant feedwater subsystem. *IEEE Transactions on Nuclear Science*, **43,** 266–277.

DINADIS, N. & VICENTE, K. J. (in press). Designing functional visualizations for aircraft system status displays. *International Journal of Aviation Psychology*.

GOODSTEIN, L. P. (1983). An integrated display set for process operators. In G. JOHANNSEN & J. E. RIJNSDORP, Eds. *Analysis, Design and Evaluation of Man–Machine Systems*, pp. 63–70. Oxford: Pergamon Press.

HOWIE, D. E., JANZEN, M. E. & VICENTE, K. J. (1996). *Research on Factors Influencing Human Cognitive Behaviour (III)* (CEL 96-06). Toronto: University of Toronto, Cognitive Engineering Laboratory.

ITOH, J., SAKUMA, A. & MONTA, K. (1995). An ecological interface for supervisory control of BWR nuclear power plants. *Control Engineering Practice*, **3,** 231–239.

MORAY, N. & ROTENBERG, I. (1989). Fault management in process control: eye movements and action. *Ergonomics*, **32,** 1319–1342.

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

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

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

VICENTE, K. J. (1991). *Supporting knowledge-based behavior through ecological interface design*. Doctoral Dissertation. Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL (unpublished).

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.

VICENTE, K. J. (1992*a*). Multilevel interfaces for power plant control rooms I: an integrative review. *Nuclear Safety*, **33,** 381–397.

VICENTE, K. J. (1992*b*). Multilevel interfaces for power plant control rooms II: a preliminary design space. *Nuclear Safety*, **33,** 543–548.

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.

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

Paper accepted for publication by Associate Editor, Dr. G. Sundstrom.